

Current status and development trend of data science analysis tools based on big data

Duo LI

Shenyang City University, Shenyang, Liaoning, 110112, China

liduosy@live.com

Keywords: big data, scientific data, analysis and discussion.

Abstract: In the current era, the application of big data has become an important basis for social development. Huge and accurate big data allows users to obtain relevant information more scientifically and accurately. Therefore, the role of big data is extremely important. This article Scientifically analyze the data in the context of big data, clarify the current specific situation and future development trends, analyze and discuss existing problems, make every effort to promote the development of big data, and make due contributions to the development of the times.

1. Introduction

Data science has gradually embodied its own meaning in the environment of big data development, using automated methods to extract useful information from a large amount of data. The so-called data science is a comprehensive data knowledge that combines numbers, computers, statistics and probability. It has a wide range of applications and can bring beneficial effects to various fields in our country. Its unique advantages allow big data users to make more correct development decisions.

2. Data Science

In the current era of big data development, a discipline called data science has emerged. At present, my country's data science is still in its infancy. Although it is still undergoing continuous development, its difficulty is also quite huge. The use of automated processing to process and analyze massive amounts of data is the main core of data science. Extracting relevant information from the data has brought a huge improvement in data exploration and discovery, and has brought new styles to many companies. Sources of insights. With the continuous improvement of automated systems, the analysis and application of data science has affected people's social life and humanities.

Data science integrates a variety of disciplines, and has higher achievements in the theoretical and technical support of various disciplines, including data visualization, statistics, probability and another related knowledge. In the actual operation process, data science also includes collection, analysis, screening, visualization, etc., bringing users a broader range of social development information, based on scientific data analysis, plus various professional experience and with knowledge, data scientists can formulate actual strategies and plans that best fit the current social development through the analysis of data. Big data analysis technology is also the greatest project of the "21st century". At present, there are still differences in the definition of data science. Many data scientists use their own perspectives and discoveries to define data science. Experts from prestigious universities at home and abroad have conducted extensive discussions and research on data science, but in theory There are still different opinions on technology and technology. Therefore, data science is also listed as an existence that is different from other disciplines. Experts and professors have different opinions. Therefore, it should also carry out more long-term practical experience to prove itself actual value.

3. The current status of data science analysis tools in the context of big data

Big data is the process of collecting statistics on the hotspot data in the current network through the Internet platform, which can assist people in collecting and analyzing the current hotspot data. The main tools for current big data analysis are as follows:

3.1 Rlanguage analysis tool

The R language analysis tool, which is an editing language and environment in the system, can be used for drawing and statistics work using the R language analysis tool. To provide people with relevant work needs, the main application methods are: (1) Based on the source code, the technology can provide people with more trial and download methods under the CNU agreement. In order to provide convenience, R language the analysis tool can provide the required CRAN comprehensive archiving technology according to the corresponding website, and provide more benefits for third-party applications. At present, the system is commonly used in social surveys, biological information and statistical analysis. Advantages make R language analysis tools widely loved by people in society.

3.2 Rlanguage analysis tools have unique compatibility

The R language analysis tool has unique compatibility. It can run normally in various computer versions, such as the widely used windows system, MACOS system and Linux system on the market. Many of the algorithms on the market can perform effective operations, such as KSS algorithm, Neural Network algorithm, EM algorithm and K-Mean's algorithm. All of these algorithms perform data mining work with technology, and have the compatibility of data formats. Excellent performance, able to process various unstructured, structured, and semi-structured data formats, and has considerable advantages in visualization, making it an outstanding contribution to graphics. Pictures in various formats are It can be drawn with R language analysis tools. Although it has considerable advantages, it also has some shortcomings. For example, in terms of running speed and secondary development, R language analysis tools have shortcomings.

3.3 The RapidMiner analysis tool has higher operability

The RapidMiner analysis tool has higher operability and is also called the YALE tool. The YALE tool has its own unique characteristics. It can perform a wide range of data analysis, and can also use JAVA API and JAVAGUI modes for collection and analysis. In the actual use process, YALE tools use data mining, machine learning, business analysis and prediction analysis methods to make environmental changes to computer systems. It has the advantage of high convenience. Due to its convenient operation and high efficiency, it is very suitable for beginners, the use of this tool is mainly as follows: (1) YALE tool has a toolbox function that suits its own use, and its function can bring greater editability to users. YALE tool can be used in data processing. Do it more quickly and conveniently. It can design the model in a very short time, and it can guarantee accuracy in terms of data mining, which greatly meets the psychology of users or customers, and further enhances the attractiveness of products. In terms of product sales Can play a huge role

3.4 The RapidMiner analysis tool has the same characteristics as the R language analysis technology

In the process of actual use by personnel, RapidMiner analysis tools and R language analysis technology have the same characteristics, can be run on different computer systems, have quite superior compatibility, and can be used in various data mining tasks. In PDF, HTML and other formats of web pages and EXCEL, ARFF and CSV and other data formats, it can bring corresponding technical support, not only has strong visualization capabilities, but also can use its own drawing functions to make pictures, the specifications of the pictures produced are more extensive, such as 3D, 2D and 1D as people know, they also have the ability to produce scatter matrix and icons, and they can effectively guarantee the running speed. Relevant data can be analyzed in five minutes, but it also has some flaws, that is, it cannot be used for second

development.

3.5 MaRout analysis tool

MaRout analysis tool, which can establish a freely scalable machine learning algorithm, and has its own database. The powerful database and algorithm capabilities can provide users with more efficient work efficiency in the analysis of big data. Its main workflow is as follows: (1) MaRout analysis tool, with the support of the two Bayesian statistical methods, can perform more efficient and scientific analysis of big data. Through the data mining of each sub-project, the use of filtering and the analysis method is used to carry out big data analysis and statistics work. (2) MaRout analysis tool also has great advantages in compatibility, whether it is MacOSX system, Linux system, Windows system or MacOSX system, it can run perfectly. In the process of use, a part of the data can be processed on the distance algorithm, which not only has an efficient running speed, but also has certain advantages in distributed data processing. Through its high-tech computing mode, it can have greater secondary development. The advantages. Although it has an irreplaceable role in secondary development, its shortcomings are also more obvious, that is, there are deficiencies in data format compatibility and visualization, and lack of drawing functions. At the same time, it cannot be effective in the face of SequenceFile format data. deal with.

4. The development trend of big data analysis

In the context of the current big data analysis, more attention should be paid to its future development trends: (1) Due to the continuous development of society, a large amount of data will appear, and these data analysis tools still lack the development of diversity and face a large number of For the generation of data, analysis tools should meet the needs of more excavator analysis, in order to obtain more valuable information data. (2) In terms of visualization, more attention should be paid. The future data analysis tools must have the conditions for visualization. The realization of visualization can make the data more intuitive, so that people can get the value of the data more quickly, smoothly and accurately. information. (3) In the next period of time, data analysis tools should be continuously developed to meet people's tasks of massive data processing, and the running speed and analysis efficiency should be comprehensively improved, so that users can obtain analysis more efficiently. As a result, data processing can be more convenient.

5. Conclusion

In combination with the article, the current status of big data tools and future development trends have been analyzed. In the future, big data analysis will be developed accordingly to achieve the comprehensiveness of tools and further visualization and diversification. Make improvements so that big data analysis tools can bring greater market effects, and bring more scientific and developmental analysis data to the market, so as to ensure that the current society and the future society can make better use of big data analysis technology stand up.

References

- [1] Jiangang Liu, Junchan Zhao, Qianhong Wan, Jianwen Chen, Yunhua Liao. Thoughts on the training of data science and big data professionals under the background of the new era [J]. Computer Age, no.02, pp.86-89, 2021.
- [2] Jie Song. Opportunities and Challenges of Science Education in the Context of Big Data—Taking the Novel Coronavirus Pneumonia as an Example[J]. Modern Primary and Secondary Education, vol.37, no.02, pp.19-23, 2021.